

# Data fusion algorithms for Density Reconstruction in Road Transportation Networks

Enrico Lovisari, Carlos Canudas de Wit, and Alain Y. Kibangou

**Abstract**—This paper addresses the problem of density reconstruction in traffic networks with heterogeneous information sources. The network is partitioned in cells in which vehicles flow from their origin to their destination. The state of the network is represented by the densities of vehicles in each cell. Density estimation is of crucial importance in future Intelligent Transportation Systems for monitoring, control, and navigation purposes. However, deploying fixed sensors for this purpose can be very expensive. Therefore, most of fixed sensors networks are rather sparse. On the contrary, recent technologies have enormously increased the availability of relatively inexpensive Floating Car Data. A data fusion algorithm is then proposed to incorporate the two sources of information into a single observer of density of vehicles. The efficiency of the proposed algorithm is shown in a real scenario using data from the Grenoble Traffic Lab fixed sensor network and INRIX Floating Car Data on the Rocade Sud in Grenoble.

**Index Terms**—Road Transportation systems; Dynamical flow network; Density reconstruction; Floating Car Data.

## I. INTRODUCTION

The increase of the number of vehicles observed in the last decades has not been matched by a comparable extension of roads infrastructure, therefore steering crucial highways and arterial roads towards a state of near saturation and exhibiting on a daily basis periods of congested traffic [1]. Congestion causes in turn increased travel times and stop-and-go phenomena, leading to decreased safety, economical losses, and environmental and psychological hazards in terms of pollution and road rage [2]. Physical constraints prevent extension or construction of highways and other arterial roads, which, by increasing road capacity, have been in the past a standard way to cope with congestion problems. Built upon recent technological and theoretical advancements, Intelligent Transportation Systems (ITSs) are expected to provide better and robust techniques for real-time monitoring, prediction and actuation of traffic networks.

This paper is devoted to studying the problem of estimating road usage in terms of density of vehicles in a traffic network. The latter is commonly considered a good representation of the state of the system, and it is of fundamental importance for 1) forecasting travel time and traffic evolution, along with historical data; 2) informing in real-time drivers about the state of the network through navigation systems; 3) providing public authorities with statistical data to monitor the state of the network and predict dangerous scenarios; 4) computing and actuating control actions through traffic

lights, ramp metering and speed limits, or, in the future, lane change and origin-destination suggestions [3], [4], [5], [6].

The main tools to estimate, or reconstruct, the state of the network are fixed sensors, such as induction loops or magnetometers. These devices are able to 1) count the number of vehicles that cross a certain section of road, 2) provide an estimate of the density of the group of vehicles that crossed the section (more precisely, their occupancy, see Section IV), and 3) measure their average speed. While the so obtained measurements usually exhibit high performance, deployment or extension of a sensing network requires considerable initial investment and maintenance. As a consequence, sensing networks are usually designed to be as sparse as possible. The problem of trading off between performance and cost is called Optimal Sensor Placement problem, and it is the focus of the companion paper [7]. In the present paper, conversely, we assume that the sensors' locations are given and fixed.

In recent times, the enormous spread of wireless devices has opened possibilities for sensing and communication unforeseeable up to few years ago. In particular for the traffic application, any vehicle equipped with devices able to compute the position and the speed of the vehicle (such as GPS location) and to communicate it to an ITS monitoring system can act as a probe in the traffic and provide Floating Car Data (FCD). In the scenario in which a non negligible fraction of vehicles accepts to act as probe, the collected data can be used to estimate of the evolution of speed in the network. Due to privacy reasons, single vehicles traces are usually not directly used, but rather aggregated as average speed of vehicles in segments of road. Advanced methodologies, such as the one employed by INRIX, ensure a very fine spatial partition of the network, with segments as short as 250 meters ([8]). Compared to fixed sensors, this technology is less precise, due to location uncertainties and to the fact that only a subset of vehicles is used to measure speed, but since it exploits existing communication systems it is relatively less expensive and, more important, already covers all major traffic networks.

In this paper we propose an algorithm that aims to reconstruct the traffic density by fusing fixed sensors measurements and Floating Car Data. We employ a macroscopic model, partitioning the network in cells and assigning to each cell a density of vehicles. The latter evolves dynamically according to a first order mass-conservation law.

Traffic models date back to the first half of the XXth century. The most celebrated macroscopic model is the PDE based Lighthill-Whitham and Richards (LWR) model [9], which, based on fluid kinematics, is able to reproduce crucial

The authors are with Univ. Grenoble Alpes, Gipsa-Lab, with CNRS, Gipsa-Lab, and with Inria Grenoble Rhône-Alpes, F-38000 Grenoble, France [enrico.lovisari](mailto:enrico.lovisari), [carlos.canudas-de-wit](mailto:carlos.canudas-de-wit), [alain.kibangou@gipsa-lab.fr](mailto:alain.kibangou@gipsa-lab.fr)

phenomena such as traffic shock waves. Discretization of the LWR-PDE is not straightforward but stable numerical schemes have been proposed, the most well known being the Cell Transmission Model (CTM) [10], [11]. Huge efforts have been put in the last 15 years to calibrate the CTM [12] and to unveil its system-theoretical properties [13]. Fusion of flow, density and speed measurements has also been addressed, even though mostly considering single vehicles traces. Approaches range from signal processing techniques such as the generalized Treiber-Helbing filter [14], nonlinear versions of the Kalman filter in the context of Lagrangian sensing [15], and stochastic versions of the three-detector model [16]. Recent approaches do not rely on discretization of the LWR-PDE model and allow to cast problems of estimation and control as convex problems [17].

We inherit from the CTM the assumption that the inflow in a cell is a fixed linear combination of the outflows of the preceding cells. Differently from CTM, however, inflows and outflows in all the cells are estimated on the basis of the available flow measurements only. In addition, using the concept of Fundamental Diagram and the speed measurements, we compute an instantaneous (namely, only based on the latest available measurements) noisy estimate of the density. These quantities are then the inputs for the density observer. Finally, we propose a gradient descent method to calibrate the Fundamental Diagram.

To summarize, the contributions of this paper are: 1) we formalize the problem of data fusion of fixed sensors measurements and Floating Car Data; 2) we propose an easily implementable approach for density reconstruction; 3) we propose a gradient descent calibration algorithm; 4) we implement the proposed solution using real fixed sensor measurements from the Grenoble Traffic Lab [18], a sensing network deployed along the Rocade Sud in Grenoble, France, and speed FCD measurements provided by INRIX, one of the most well known traffic solutions companies.

The remainder of the paper is organized as follows: Section II formulates the problem and details the available measurements. Section III describes the proposed solution, while Section IV presents our numerical experiments. Finally, Section V draws the conclusions and presents several future research directions.

### A. Notation

The symbols  $\mathbb{R}^n$ ,  $\mathbb{R}_+^n$  and  $\mathbb{R}^{n \times m}$  denote the sets of real valued vectors of dimension  $n$ , of positive real valued vectors of dimension  $n$ , and of real valued matrices of dimension  $n \times m$ , respectively. The symbol  $\mathbb{R}^{\mathcal{A}}$  with  $\mathcal{A}$  a finite set is to be interpreted as the set of vectors indexed by elements of  $\mathcal{A}$ . The transpose of  $A \in \mathbb{R}^{n \times m}$  is denoted  $A^T$ . For a vector  $x \in \mathbb{R}^n$ ,  $x \geq 0$  is meant component-wise. The symbol  $I$  denotes the identity matrix of suitable dimensions.  $|\mathcal{A}|$  denotes the cardinality of the set  $\mathcal{A}$ .

A graph  $\mathcal{G}$  is a pair  $(\mathcal{V}, \mathcal{E})$  where  $\mathcal{V}$  is called the set of nodes and  $\mathcal{E}$  the set of edges. Edges are equipped with two functions  $t : \mathcal{E} \rightarrow \mathcal{V}$  and  $h : \mathcal{E} \rightarrow \mathcal{V}$ , the head and tail functions, respectively, so that  $e$  is the directed edge between

$t(e)$  and  $h(e)$ . We allow for parallel edges, namely, edges having the same head and tail, but not for loops, namely, edges whose head and tail coincide. A path of length  $n \geq 2$  is a sequence of edges  $e_1, \dots, e_n$  that are consecutive, namely, such that  $h(e_i) = t(e_{i+1})$  for  $i = 1, \dots, n-1$ . A path of length 1 is a path made of a single link.

## II. PROBLEM FORMULATION

We model a Transportation Network as a graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , in which junctions  $v \in \mathcal{V}$  are interfaces between cells  $e \in \mathcal{E}$ . Roads with more than one lane are partitioned in sequences of parallel cells on the parallel lanes; onramps and offramps, namely, roads carrying vehicles in and out from the network, respectively, are considered cells as well. For physical reasons, we assume that for each cell  $e$  there is at least one onramp  $j$  and one offramp  $k$  such that  $e$  is on a path from  $j$  to  $k$ . Time is discrete and is slotted in intervals of length  $T > 0$ . On each cell  $e \in \mathcal{E}$  we denote by  $\rho_e(t)$  the density of vehicles, in number of vehicles per km, during the  $t$ -th time slot. The vector of densities  $\rho(t) = [\rho_e(t)]_{e \in \mathcal{E}}$ , the *state* of the network, changes dynamically in time according to the following first-order equation model

$$\rho_e(t+1) = \rho_e(t) + \frac{1}{\ell_e} (f_e^{\text{in}}(t) - f_e^{\text{out}}(t)) \quad (1)$$

where  $\ell_e$  is the length of cell  $e$ , and  $f_e^{\text{in}}(t)$  and  $f_e^{\text{out}}(t)$  are the inflow and the outflow at cell  $e$  during the  $t$ -th time slot.

The fraction of vehicles on cell  $e$  that will turn into the cell  $j$  is called the splitting ratio of the pair  $(e, k)$  and is denoted  $R_{ej} \geq 0$ . If  $j$  does not physically follow  $e$ , then  $R_{je} = 0$ , and moreover  $\sum_{j \in \mathcal{E}} R_{ej} \leq 1$ , with strict inequality at offramps. Using the splitting ratios, the inflow in a cell  $e$  is a function of the outflows from the preceding cells,  $f_e^{\text{in}} = \sum_{j \in \mathcal{E}} R_{je} f_j^{\text{out}}$ . Stacking inflows and outflows into vectors  $f^{\text{in}}$  and  $f^{\text{out}}$ , we can rewrite the previous relation in matrix form as

$$f^{\text{in}} = R^T f^{\text{out}} \quad (2)$$

where the matrix  $R = [R_{ej}]$  is the *matrix of splitting ratios*. In this paper we assume that the matrix of splitting ratios is fixed and predetermined. Methods for its calibration, which is closely related to the estimation of Origin-Destination pairs, will be the focus of future research.

Well known macroscopic models such as the CTM postulate that the outflow  $f_e^{\text{out}}$  on cell  $e$  depends on the density of vehicles on cell  $e$  *as well as* on the density of other cells, including those immediately following cell  $e$ . This assumption descends by the Godunov scheme for discretization of the LWR-PDE model, and expresses the fact that while vehicles on cell  $e$  want to proceed further into subsequent cells, the latter could be too congested to let more vehicles in.

While these models can reproduce important phenomena, such as shockwaves, there is no universal agreement on which is the best. For this reason, we adopt a different point of view and renounce to explicitly determine the flow-density relation. The unique assumption that we make is the standard

$$f_e^{\text{out}} = \rho_e v_e(\rho), \forall e \in \mathcal{E} \quad (3)$$

namely that the outflow from a cell is the product of density and average speed of vehicles on the same cell. While in the CTM the speed is a function of the density of vehicles around cell  $e$ , we will leave it unmodelled. The reason is that, as it will be clarified in the following, we assume having measurements of speed in each cell, so there is no need to model its dependence on the state of the network. We shall consider from now on the dynamics of the real system to be dictated by Eqs. (1)-(3), where  $v_e(\rho)$ ,  $e \in \mathcal{E}$ , is an unmodelled function of the local state of the network.

For data fusion and estimation purposes, we make a further simplifying assumption, writing

$$\varphi_e = \varphi_e(\rho_e), \forall e \in \mathcal{E} \quad (4)$$

where  $\varphi_e$  is the flow of vehicles *at the sensor locations*. The graph of the function  $\varphi_e(\cdot)$  is called the Fundamental Diagram on cell  $e$ , and it is usually supposed to be a concave function with  $\varphi(0) = \varphi(\rho^{\text{jam}}) = 0$ , where  $\rho^{\text{jam}}$  is the jam, or stopping, density (standard values for  $\rho^{\text{jam}}$  vary from 200 to 300 vehicles per km).

#### A. Available measurements

As already mentioned, in this paper we consider a scenario in which heterogeneous measurements are available.

1) *Flow and density measurements*: Measurements of flows and of density are obtained by fixed sensors, such as classical loop detectors, magnetometers, radar traffic detectors, or video detection systems. These devices count the number (flow) and estimate the density of the vehicles that, during a time slot, cross the section of road where they are placed. Sources of noise range from temporary inability to detect changes of the magnetic field, too fast or too slow vehicles, blurred videos, etc. For sake of simplicity, we assume from now on that a new set of flow and density measurements is available at every time slot. This is done without any loss of generality and for the sole sake of simplifying the notation. We can thus write

$$\begin{aligned} \varphi_e^m(t) &= f_e^{\text{out}}(t) + \omega_e^\varphi(t), & e \in \mathcal{E}^m \\ \rho_e^m(t) &= \rho_e(t) + \omega_e^\rho(t), & e \in \mathcal{E}^m \end{aligned} \quad (5)$$

where  $\varphi_e^m(t)$  and  $\rho_e^m(t)$  are flow and density estimates at time  $t$ , and  $\omega_e^\varphi(t)$  and  $\omega_e^\rho(t)$  are measurement errors whose stochastic properties depend on the performance of the sensor as well as on road and weather conditions, and  $\mathcal{E}^m \subseteq \mathcal{E}$  is the set of cells equipped with sensors. Due to installation and maintenance costs, usually  $|\mathcal{E}^m| \ll |\mathcal{E}|$ .

2) *Speed measurements*: Vehicles embedded with communication and tracking devices (navigation systems, smartphones, etc) can communicate their speed to public or private ITS monitoring systems in the form of Floating Car Data, including their position, speed and travel direction (traces). Since for privacy reasons traces of private vehicles cannot be stored and used, data are then aggregated in average speed. In particular, roads are divided into segments, let  $\mathcal{S}$  denote the set of all segments, and one average speed is provided for each segment  $s \in \mathcal{S}$ . Notice that the resulting measurement is an average of the speed of a *fraction* of vehicles in the road.

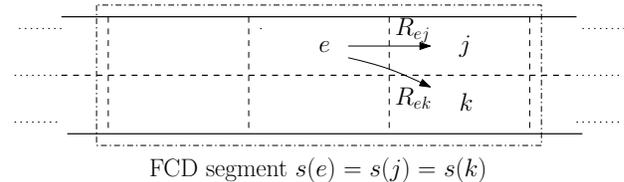


Fig. 1. A stretch of road partitioned in cells and FCD segments. Splitting ratios are shown from a cell  $e$  to following cells,  $j$  and  $k$ . A FCD segment including, among others, cells  $e$ ,  $j$  and  $k$ , is also shown.

While there are specific classes of vehicles, such as taxis and buses, that do not have privacy requirements, but we will not consider them in the present paper, leaving the possibility to exploit this additional information in future research.

Floating Car Data are usually much less expensive than fixed sensors as they leverage on the existing communication architecture, which also make them available on great part of the main freeways and roads. As drawbacks, their precision is not as high as fixed sensors, being related to the penetration rate of the ITS provider, and they do not distinguish among vehicles in different lanes. Finally, the information provided via FCD is usually averaged over a relatively long period of time. For example, while the GTL sensor network provides flow, density, and speed measurement every 15 seconds, FCDs are aggregated by INRIX every minute, standard practice being around 5/10 minutes. We will assume that new speed aggregate data is available every  $N$  time instants, i.e., at times  $N, 2N, 3N, \dots$ , corresponding to the average speed in the periods  $[0, N-1]$ ,  $[N, 2N-1]$ , etc., respectively. As such, speed measurements can be formally written as

$$v_e^{\text{FCD}}(t) = \begin{cases} v_e^{\text{ff}}, & t \in [0, N-1] \\ v_{s(e)}^{\text{FCD}}(k), & t \in [kN, (k+1)N-1] \end{cases} \quad (6)$$

where

- $v_e^{\text{ff}} > 0$  is the freeflow speed on cell  $e$ , namely, the speed of vehicles in low density regime;
- $v_{s(e)}^{\text{FCD}}(k)$  is given by

$$v_{s(e)}^{\text{FCD}}(k) = \frac{1}{N|s(e)|} \sum_{j \in s(e), \tau \in \mathcal{I}_t} v_j(\tau) + \omega_{s(e)}^{\text{FCD}}(k)$$

where  $s(e)$  denotes the segment of which  $e$  is one of the cells (see Figure 1),  $\omega_{s(e)}^{\text{FCD}}(k)$  is a measurement error whose stochastic properties depend on the performance of the sensor as well as on road and weather conditions, and  $\mathcal{I}_t = \{\tau : \lfloor \frac{t}{N} \rfloor - 1 \leq \frac{\tau}{N} < \lfloor \frac{t}{N} \rfloor\}$ .

### III. A NONLINEAR OBSERVER FOR TRAFFIC NETWORKS

We observe that Eqs. (1)-(2) could be directly used to observe the system only in the ideal scenario in which the traffic monitoring system has precise measurements of the outflows  $f_e^{\text{out}}(t)$ , for all  $e \in \mathcal{E}$  and for all times  $t \geq 0$ , and of the initial conditions of the system. Real systems are never error free, however, and it is clear from Eq. (1) that while errors in the initial conditions remain as offsets during the evolution of the system, errors in the flow measurements enter in the dynamics integrated, therefore

possibly producing unbounded estimation errors. Therefore, Eqs. (1)-(2) cannot be used as they are to observe the system.

To overcome this difficulty, we consider the following standard Luenberger-like observer, in which  $\varphi^m(t) = [\varphi_e^m(t)]_{e \in \mathcal{E}^m}$  and  $v^{\text{FCD}}(t) = [v_e^{\text{FCD}}(t)]_{e \in \mathcal{E}}$  are the stacked versions of the above described measurements,

$$\begin{cases} \hat{\rho}_e(t+1) = \hat{\rho}_e(t) + \frac{1}{\ell_e}(\hat{f}_e^{\text{in}}(t) - f_e^{\text{out}}(t)) \\ \quad + \kappa(\tilde{\rho}_e(t) - \hat{\rho}_e(t)) \\ \hat{f}_e^{\text{in}}(t) = \hat{f}_e^{\text{in}}(\varphi^m(t)) \\ \hat{f}_e^{\text{out}}(t) = \hat{f}_e^{\text{out}}(\varphi^m(t)) \\ \tilde{\rho}_e(t) = \tilde{\rho}_e(\varphi^m(t), v^m(t)) \end{cases} \quad \forall e \in \mathcal{E} \quad (7)$$

where

- $\hat{\rho}_e(t)$  is the estimate of the density on cell  $e$  at time  $t$ ;
- $\hat{f}_e^{\text{in}}(t)$ ,  $\hat{f}_e^{\text{out}}(t)$  are estimates, based on the flow measurements, of inflow and outflow in cell  $e$  at time  $t$ ;
- $\tilde{\rho}_e(t)$  is an instantaneous estimate, based on flow and speed measurements, of the density on cell  $e$  at time  $t$ ;
- $\kappa$  is a tunable gain trading off between flow and instantaneous density estimates. Notice that it should not be too large, as to avoid meaningless negative densities.

The problem that we tackle in this paper is how to design the maps

$$\begin{aligned} \hat{f}^{\text{in}} &= \{\hat{f}_e^{\text{in}}\}_{e \in \mathcal{E}} : \mathbb{R}_+^{\mathcal{E}^m} \rightarrow \mathbb{R}^{\mathcal{E}} \\ \hat{f}^{\text{out}} &= \{\hat{f}_e^{\text{out}}\}_{e \in \mathcal{E}} : \mathbb{R}_+^{\mathcal{E}^m} \rightarrow \mathbb{R}^{\mathcal{E}} \\ \tilde{\rho} &= \{\tilde{\rho}_e\}_{e \in \mathcal{E}} : \mathbb{R}_+^{\mathcal{E}^m} \times \mathbb{R}_+^{\mathcal{E}} \rightarrow \mathbb{R}^{\mathcal{E}} \end{aligned}$$

in such a way that the observer provides a good estimate of the real density of the system.

### A. Proposed solution

This section is devoted to describing the proposed solution. The procedure consists of two steps, namely offline calibration and online update.

1) *Offline calibration*: In this section we propose a simple solution for calibrating the Fundamental Diagram, namely, for estimating the function  $\varphi_e(\cdot)$  on the cells  $e \in \mathcal{E}^m$ . For cells in  $\mathcal{E} \setminus \mathcal{E}^m$  the Fundamental Diagram can be then obtained by extending the parameters by convex interpolation.

In this paper we consider a the following type of Fundamental Diagram

$$\varphi_e(\rho) = \begin{cases} v_e^{\text{ff}} \rho, & \rho \leq \rho_e^c \\ a_e \rho^2 + b_e \rho + c_e, & \rho > \rho_e^c \end{cases}$$

where

- $\rho_e^c$  is the critic density, which separates the freeflow low-density region  $[0, \rho_e^c)$ , in which vehicles do not influence one each other, from the high-density congested region  $(\rho_e^c, \rho_e^{\text{jam}}]$ , in which speed decreases with density;
- $\rho_e^{\text{jam}}$  is the jam density, at which vehicles stop;
- $v_e^{\text{ff}} > 0$  is the freeflow speed on cell  $e$ ;
- the flow in congested region is a convex quadratic function of the density, and for consistency the parameters

$a_e$ ,  $b_e$  and  $c_e$  satisfy

$$\begin{cases} a_e (\rho_e^c)^2 + b_e \rho_e^c + c_e = v_e^{\text{ff}} \rho_e^c \\ a_e (\rho_e^{\text{jam}})^2 + b_e \rho_e^{\text{jam}} + c_e = 0 \\ a_e > 0 \end{cases}$$

The previous choice is driven by the empirical observation that the standard triangular diagram, in which the flow is a piecewise linear affine function of the density, tends to overestimate the flow in congestion. Alternative solutions include employing the so called inverted- $\lambda$  fundamental diagram [19], but in this case the number of parameters to be estimated grows and the resulting model becomes more complex, involving hysteresis. Since, in any case, deterministic Fundamental Diagrams are rough representations of the relation between flow and density (see Section IV), we chose the quadratic function because it allows for reasonably good performance despite being relatively simple to calibrate.

Let  $e \in \mathcal{E}^m$  be a cell equipped with sensors, and denote by  $\{(\rho_k, \varphi_k)\}_{k \in \mathcal{K}}$ ,  $\mathcal{K} = 1, \dots, K$ , the set of  $K$  density and flow measurements used as learning set. Calibration of the Fundamental Diagram involves two steps (we write  $\rho^c$  and  $C$  instead of  $\rho_e^c$  and  $C_e$  for sake of notation)

- Gradient descent algorithm for estimation of  $\rho^c$  and  $C = v^{\text{ff}} \rho^c$ : in the first step, we estimate the critical density  $\rho^c$  and the capacity  $C$ , namely, the *nominal* maximum flow, by solving the non-linear and non-convex minimization problem

$$\begin{aligned} \min_{(\rho^c, C)} \quad & V_{(\rho^c, C)} = \sum_k (\varphi_k - \varphi_{(\rho^c, C)}(\rho_k))^2 \\ \text{s.t.} \quad & 0 < \rho^c < \rho^{\text{jam}} \\ & C > 0 \\ & \varphi_{(\rho^c, C)}(x) = \begin{cases} \frac{C}{\rho^c} x, & x \leq \rho^c \\ \frac{C(\rho^{\text{jam}} - x)}{\rho^{\text{jam}} - \rho^c}, & x > \rho^c \end{cases} \end{aligned} \quad (8)$$

To solve (8) we propose the following gradient descent with diminishing stepsize algorithm

- Basic step: initialize  $\rho_0^c, C_0$ . A possible choice is  $\rho_0^c = 20$  (vehicles start influencing each other when their relative distance is less than 50 meters), and  $C_0 = v_e^{\text{limit}} \rho_0^c$ , where  $v_e^{\text{limit}}$  is the speed limit on cell  $e$  normalized by the sampling time  $T$ ;
- $n$ -th step: let  $(\rho_n^c, C_n)$  descend along the gradient of the cost, namely

$$\begin{aligned} \rho_{n+1}^c &= \rho_n^c + \frac{\delta}{n} \nabla_{\rho^c} V_{(\rho^c, C)} \\ C_{n+1} &= C_n + \frac{\delta}{n} \nabla_C V_{(\rho^c, C)} \end{aligned}$$

with

$$\begin{aligned} \nabla_{\rho^c} V_{(\rho^c, C)} &= \sum_{k \in \mathcal{I}^{\text{FF}}(\rho_n^c)} -\frac{C_{n-1}}{(\rho_{n-1}^c)^2} \rho_k \\ &+ \sum_{k \notin \mathcal{I}^{\text{FF}}(\rho_n^c)} C_{n-1} \frac{\rho^{\text{jam}} - \rho_k}{(\rho^{\text{jam}} - \rho_{n-1}^c)^2} \end{aligned}$$

$$\begin{aligned}\nabla_C V_{(\rho^c, C)} &= \sum_{k \in \mathcal{I}^{\text{FF}}(\rho_n^c)} \frac{\rho_k}{\rho_{n-1}^c} \\ &+ \sum_{k \notin \mathcal{I}^{\text{FF}}(\rho_n^c)} \frac{\rho^{\text{jam}} - \rho_k}{\rho^{\text{jam}} - \rho_{n-1}^c} \\ \mathcal{I}^{\text{FF}}(\rho^c) &= \{k \in \mathcal{K} : \rho_k \leq \rho^c\}\end{aligned}$$

where the gradients  $\nabla_{\rho^c} V_{(\rho^c, C)}$  and  $\nabla_C V_{(\rho^c, C)}$  are computed at  $(\rho^c, C) = (\rho_{n-1}^c, C_{n-1})$ , and  $\delta > 0$  is a fixed initial step size;

- Stopping criterion: stop if  $\| \begin{bmatrix} \rho_n^c \\ C_n \end{bmatrix} - \begin{bmatrix} \rho_{n-1}^c \\ C_{n-1} \end{bmatrix} \| < \varepsilon$  for some small threshold  $\varepsilon > 0$ .

- Calibration of the congested part: once  $(\rho^c, C)$  have been estimated, the problem of calibrating the quadratic function for the congested region can be cast as the following quadratic problem

$$\begin{aligned}\min_{(a,b,c)} & \sum_{k \in \mathcal{I}^{\text{FF}}(\rho^c)} (\varphi_k - (a\rho_k^2 + b\rho_k + c))^2 \\ \text{s.t.} & a(\rho^c)^2 + b\rho^c + c = C \\ & a(\rho^{\text{jam}})^2 + b\rho^{\text{jam}} + c = 0 \\ & a \geq 0\end{aligned}\tag{9}$$

where the last constraint ensures that the quadratic function is convex. The problem (9) can be readily solved using off-the-shelf tools.

Once calibration has been performed on cells  $e \in \mathcal{E}^m$ , the Fundamental Diagrams can be extended on cells  $e \in \mathcal{E} \setminus \mathcal{E}^m$  by convex interpolation of the parameters  $\rho^c$ ,  $C$ ,  $a$ ,  $b$  and  $c$ , using the corresponding parameters calibrated in the closest cells  $e \in \mathcal{E}^m$ , with weights proportional to the relative distance among the cells.

2) *Online density reconstruction algorithm*: Assume that Fundamental Diagrams have been calibrated or extended on all cells  $e \in \mathcal{E}$ , and that the matrix of splitting ratios has been pre-specified or estimated on the basis of field surveys.

The algorithm proceeds as follows

- at the beginning of the  $t$ -th time slot
  - receive the measurements  $\{\varphi_e^m(t)\}_{e \in \mathcal{E}^m}$ . The vector of estimate of the outflows  $\hat{f}_e^{\text{out}}(t)$  is then computed as the solution of the following minimization problem

$$\begin{aligned}\min_{\hat{f}_e^{\text{out}}} & \|(I - R^T)\hat{f}^{\text{out}}\|^2 \\ & + \gamma \sum_{e \in \mathcal{E}^m} (\hat{f}_e^{\text{out}} - \varphi_e^m(t))^2 \\ \text{s.t.} & \hat{f}_e^{\text{out}} \geq 0\end{aligned}\tag{10}$$

which aims, on the one side, to match outflows and measurements, and, on the other side, to balance outflows according to the splitting ratios, as if the network were at steady state. The parameter  $\gamma$  trades off between the two: the higher, the more flows on cells in  $\mathcal{E}^m$  will be forced to match the measurements, the lower, the more the solution will be as if the system were at steady state. The vector of estimate of the inflows is finally computed following Eq. 2,  $\hat{f}^{\text{in}}(t) = R^T \hat{f}^{\text{out}}(t)$ .

- receive the measurements  $\{v_e^m(t)\}_{e \in \mathcal{E}}$  when available, or hold the previous measurements;
- For each cell  $e$ , assume that the outflow can be approximated by the local flow  $\varphi_e = \varphi_e(\rho_e)$  determined by the Fundamental Diagram. As such, compute the two possible densities  $\rho_e^1$  (freeflow) and  $\rho_e^2$  (congested) corresponding to flow  $\hat{f}_e^{\text{out}}(t)$ ;
- For each cell  $e$ , compute the two corresponding velocities  $v_e(\rho_e^1) = \frac{\hat{f}_e^{\text{out}}}{\rho_e^1}$  and  $v_e(\rho_e^2) = \frac{\hat{f}_e^{\text{out}}}{\rho_e^2}$ ;
- Set

$$\tilde{\rho}_e(t) = \arg \min_{i=1,2} \{|v_e(\rho_e^i) - v_e^m(t)|\}$$

This is a first estimate of the density, only based on latest measurements of flows and speed, and thus very noisy, especially in congestion regime. For this reason, we don't use it directly;

- For each cell  $e \in \mathcal{E}$ , let the density estimate evolve according to the observer equation (7).

#### IV. DENSITY RECONSTRUCTION - EXPERIMENTAL RESULTS

Our experimental setting is the Grenoble Traffic Lab (GTL) [18], a network of sensors deployed for monitoring and research purposes along the ‘‘Rocade Sud’’, a peri-urban 12 km long freeway connecting the two highways A41 (north-west) to A480 (south-east) in the town of Grenoble in the south of France, see Figure 2, upper panel. The network consists of 135 magnetometers buried in the asphalt along the main line of the freeway, on both lanes every 250 meters (on average), on each onramp and offramp, and on three queues carrying vehicles from the urban network to three onramps, totalling 68 *sensing locations*. Figure 2, upper panel, shows the position of each of the 22 sections of the main line in which there are sensing locations on both slow and fast lanes (and usually a ramp). Except for the queues and one onramp, each sensing location consists of two magnetometers deployed in pairs at a fixed distance of  $d = 2.5m$ . On each sensing location and every  $T = 15$  seconds, the system counts the number  $\varphi_e^m$  of vehicles that crossed the location, their average speed  $v_e^m$ , and the average occupancy  $o_e^m$  of the location, defined as the percentage of the last period of  $T = 15$  seconds a vehicle was sitting over the sensors. The latter is approximatively related to the density of vehicles by the relation  $\rho_e \approx \frac{o_e}{100\ell_{\text{ave}}}$ , where  $\ell_{\text{ave}}$  is the average length of a vehicle (in km). For this reason, we shall assume that sensors are able to measure density directly. Figure 2, lower panel, shows a stylized representation of the Rocade Sud, including ramps and queues, the positions of the 68 sensing locations, and the distance between consecutive measurement sections along the main line. For a detailed report on the GTL, we refer to [18].

We partition the network in cells in such a way that each cell comprises the space between two consecutive sensing locations. As such, the numbered circles in Figure 2, lower panel, also represent cells. In addition to fixed sensors, we will use Floating Car Data provided us by INRIX, one of

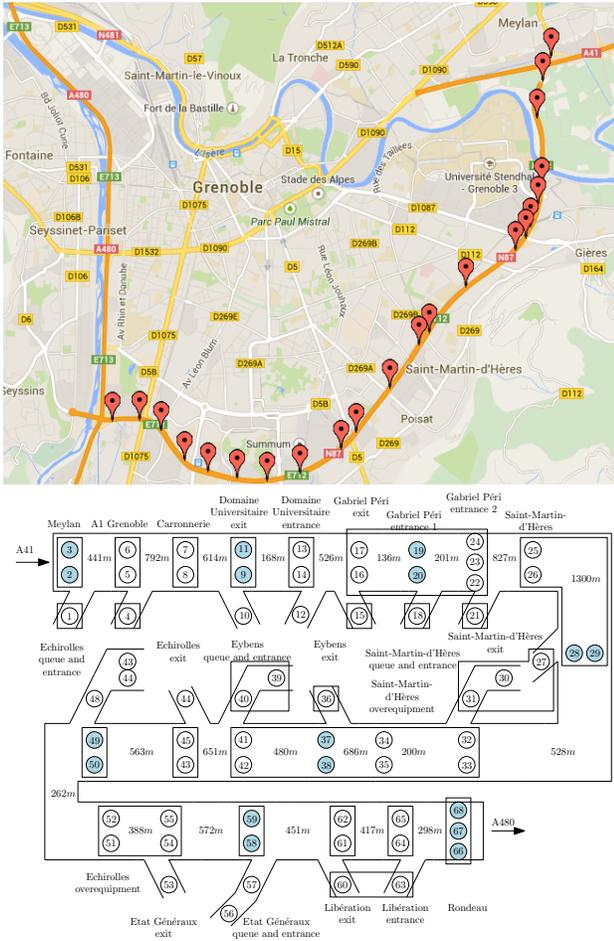


Fig. 2. The experimental setting: the town of Grenoble and the Rocade Sud (upper panel), and a stylized version of the freeway (lower panel). The positions along the main line of the 22 sections of the main line in which sensors have been placed is shown as red pin. The positions of the 68 fixed sensing locations are shown in the stylized map. Each sensing location also corresponds to a cell. Light blue circles denote fixed sensors that are used in our implementation. Each rectangles represents one FCD segment, often providing average speed measurements over more than one cell.

the main navigation and traffic monitoring companies. The Rocade has been partitioned into several *FCD segments*, and, by gathering speed information from vehicles travelling along the freeway, the average speed is computed in each FCD segment every 1 minute. FCD segments cover the whole main line of the Rocade and most onramps and offramps. Lanes are not distinguished along the main line due to the inherent uncertainty in localizing vehicles using tools such as GPS. FCD segments are represented in Figure 2 as rectangles encircling several sensing locations/cells.

For our experiments, in order to prove that the method shows good performance even with sparse equipment, we decided to consider a very limited subset of all sensors available in the network. In particular, we only use the sensors on the 9 sections shown in light blue in Figure 2, of which 8 correspond to sections of the Rocade in which loops inductors have been deployed by the Government Agency *Centre national d'information routière (CNIR)* [20] for monitoring purposes. As a consequence, we don't use

any information on flow or speed on the ramps.

We used for calibration of the Fundamental Diagram the data from April 10th, 2014, a working day (a Thursday) exhibiting very standard traffic pattern:

- very limited night time traffic;
- a peak of congestion in the morning (8:00 - 10:00), triggered by vehicles entering in the city from the Rocade at the offramp of Eybens (sensing locations 37/38) and spilling back until Meylan, and a second, smaller peak of congestion triggered by vehicles entering in A480 at Rondeau (unable to do so due to the high traffic on A480), and spilling back until around Libération;
- a third, smaller, congestion triggered around Eybens around 14:00-15:00;
- in general, medium/heavy but fluid traffic from 10:00 to 16:00
- a second peak of congestion in the afternoon, again triggered by congestion at Rondeau at around 16:00, spilling back on the whole freeway in around 60 minutes, and lasting around two hours.

The matrix of splitting ratios is set as follows:

- let  $e$  be a fast lane cell, and  $j$  and  $k$  be the following fast and slow lane cells. Then  $R_{ej} = R_{ek} = 0.5$ ;
- let  $e$  be a slow lane cell, and  $j$  and  $k$  be the following fast and slow lane cells. If among the cells that follow  $e$  there is not an offramp, then  $R_{ej} = R_{ek} = 0.5$ . Otherwise,  $R_{ej} = R_{ek} = 0.4$  and  $R_{er} = 0.2$ , where  $r$  is the offramp cell that follows  $e$ ;
- if  $e$  is an onramp cell and  $j$  is the following slow ramp cell, then  $R_{ej} = 1$ .

In words, vehicles split uniformly in the cells on the main line, and at each offramp approximately 10% of vehicles exit from the freeway, while the remaining 90% continue on the main line.

#### A. Validation

To validate our method, we chose another Thursday working day, April 24th 2014. A typical result of calibration of the Fundamental Diagram at sensing locations is illustrated in Figure 3, which shows in thick black the linear-convex Fundamental Diagram, in dashed thick black the corresponding standard linear Fundamental Diagram in congestion regime, and as crosses the pairs (density, flow) measured on April 24th, 2014. As standard and well known, data in freeflow regime are in good accordance with the linear part, while data in congested regime are much more scattered. As it can be observed, a standard bilinear Fundamental Diagram would overestimate the flows in congested regime (dashed thick line), while the convex quadratic curve seems to better capture the average flow-density relation. Nonetheless, it remains a very crude approximation of such a relation, for which a stochastic description seems much more suitable. Investigation of the latter possibility will be the focus of future research.

We implemented the proposed algorithms in Matlab on a non dedicated commercial laptop with 2.1 GHz i7-4600U

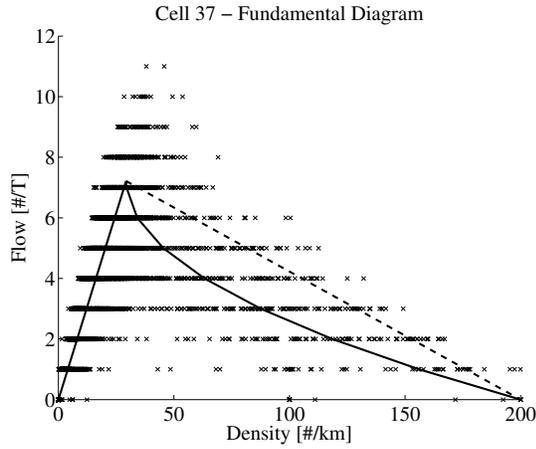


Fig. 3. Calibration of the Fundamental Diagram on the cell Eybens exit - slow lane. The linear-convex Fundamental Diagram, calibrated using data from April 10th, 2014, is shown in thick line. Dashed thick line represents the corresponding linear Fundamental Diagram in congested regime. Each cross is a (flow, density) pair measured on April 24th, 2014, one for each time slot of  $T = 15$  seconds during the whole day. Flow is the number of counted vehicles crossing the sensing location, density is the measured density of vehicles, during the time slot.

CPU and 8 GB RAM. The optimization problems whose solution is required for calibration and density reconstruction are solved using standard Matlab functions and the modelling and optimization system CVX [21], [22]. Calibration time of the Fundamental Diagram is between 30 and 40 seconds, and reconstruction of the whole day is done in around 6.30 minutes, for an average of around 1.1 ms per time slot.

The results are reported in Figure 4. For validation purposes only, density measurements from all GTL fixed sensors are considered ground truth. As such, the upper panel shows the evolution of the “true” measured density in all the cells on the main line, over the whole day. On the  $x$ -axis, the 46 sensing locations along the main line (numbers correspond to the labels in Figure 2), on the  $y$ -axis, the 5760 time slots over the whole validation day (one slot every  $T = 15$  seconds). Colors vary from green to yellow to red as density increases, with a minimum of 0 vehicles per km (green) to a maximum of  $\rho^{\text{jam}} = 200$  vehicles per km (red). In the lower panel, we show, in the same format, the results of the density reconstruction. As it can be observed, the estimation algorithm captures the four congestions described above in a reasonably good way, given the limited amount of information employed; in particular, one can appreciate the fact that the two small congestions at Rondeau during the morning and at Eybens during early afternoon are both detected.

The performance of the algorithm is further illustrated in Figure 5, which shows the percentage of pairs (time, cell) ( $x$ -axis) in which the relative error between measured and estimated densities is less than  $\delta\%$  ( $y$ -axis),

$$r(t, e) = 100 \left| \frac{\rho_e(t) - \hat{\rho}_e(t)}{\hat{\rho}_e(t)} \right| \leq \delta$$

It can be seen that 90% and 95% of (time, cell) pairs show

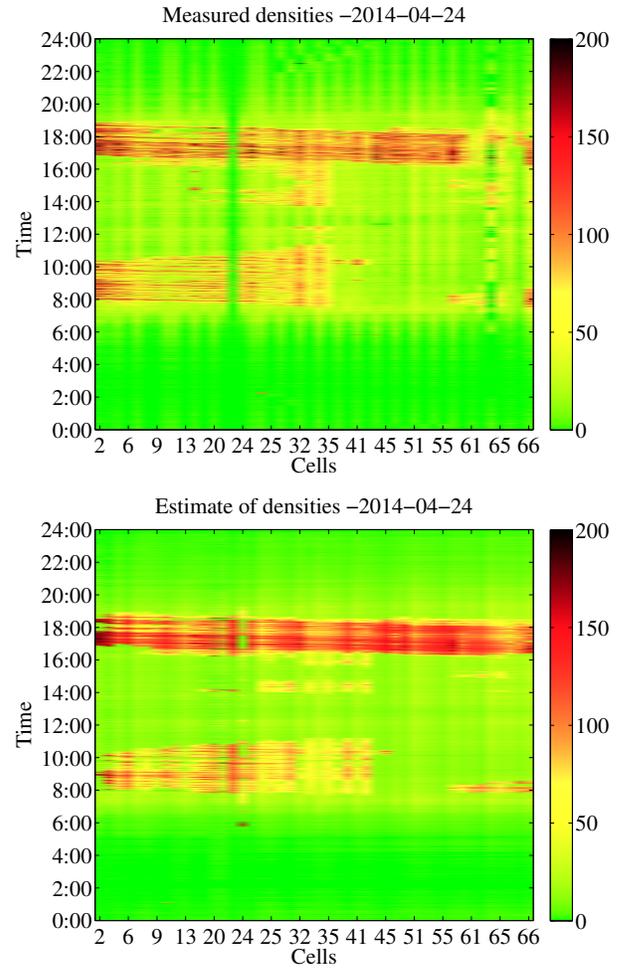


Fig. 4. Numerical results of density estimation on all cells: measured densities (upper panel) and estimated densities (lower panel).

a relative error between measured and estimated densities which is less than 1.7% and 3.5% respectively, and only less than 2% of the pairs have a relative error larger than 14.2%. We can therefore conclude that the estimated densities capture in large part the evolution of the measured densities.

The biggest discrepancy between estimates and measured data is around Eybens (cells 37-38) and can be easily explained: as mentioned, we avoid using ramp data in order to show the robustness of the method and the possibility to employ a very limited number of sensors. However, as observed above, the exit of Eybens is a critical point at which many vehicles exit the Rode, and that point belongs to a long FCD segment running from Saint-Martin-d’Hères (cells 32-33) to Eybens entrance (cells 41-42), thus providing a unique, rather low speed measurement, not distinguishing between the stretches of road before Eybens exit (congested, low speed) and after Eybens exit (uncongested, high speed). The low speed measurement causes the algorithm to believe that the density of vehicles is high over the whole segment. A second obvious discrepancy is the smoothness of the reconstructed densities in congested regime compared with more scattered density measurements. The latter can be

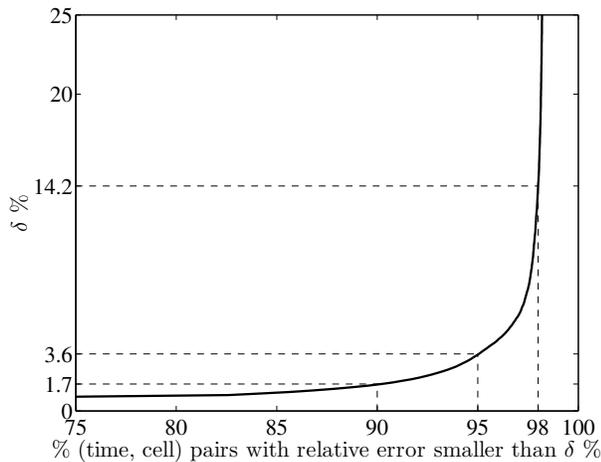


Fig. 5. Numerical results of density estimation on all cells: relative error between measured and estimated densities.

explained with the high rate of measurement of densities, implying that the state of single cells rapidly oscillates between stopped vehicles and low, but non negligible, speed (stop-and-go phenomena).

## V. CONCLUSIONS

In this paper we have formalized the problem of data fusion of heterogeneous sources of information for density reconstruction, and we have proposed an easily implementable solution that employs sparse fixed sensor measurements of flow and density and ubiquitous average speed measurements computed on the basis of Floating Car Data. Calibration algorithms for the Fundamental Diagram are also discussed. Future research directions include and are not limited to estimation of statistical properties of measurement noises from real data and development of stochastic models for the relation between flows, speed and densities (possibly on different cells) aiming to design Kalman-like filters for minimization of mean-square reconstruction error, calibration of the matrix of splitting ratios, and extension of the considered numerical scenario to part of the town of Grenoble.

## VI. ACKNOWLEDGEMENTS

The authors wish to gratefully thank Rene Fritz, Mike Corlett, and INRIX Europe, whose collaboration and permission to use data from the Grenoble area led to the results presented in this work.

## REFERENCES

- [1] M. Papageorgiou, M. Ben-Akiva, J. Bottom, P.H.L. Bovy, S. P. Hoogendoorn, N. B. Hounsell, A. Kotsialos, and M. McDonald. ITS and traffic management. In C. Barnhart and G. Laporte, editors, *Handbooks in Operations Research and Management Science: Transportation*, chapter 11, pages 715–774. Elsevier, 2007.
- [2] J. Bilbao-Ubillos. The costs of urban congestion: estimation of welfare losses arising from congestion on cross-town link roads. *Transport. Res. A-Pol.*, 42(8):1098–1108, 2008.
- [3] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Wang Y. Review of road traffic control strategies. *Proc. of the IEEE*, 91(12):2043–2067, 2003.
- [4] M. Papageorgiou, H. Hadj-Salem, and J.-M. Blosseville. ALINEA: A local feedback control law for on-ramp metering. *Transportation Research Record*, (1320):58–64, 1991.
- [5] D. Pisarski and C. Canudas de Wit. Optimal balancing of road traffic density distributions for the cell transmission model. In *Proc. of the 51th IEEE Conference on Decision and Control (CDC'12)*, page 6969 6974, Maui, Hawaii, USA, 2012.
- [6] G. Como, E. Lovisari, and K. Savla. Throughput optimal distributed routing in dynamical flow networks. In *Proc. of the 52st IEEE Conference on Decision and Control (CDC'13)*, Florence, Italy, 2013.
- [7] E. Lovisari, C. Canudas de Wit, and A. Y. Kibangou. Optimal sensor placement in transportation networks using virtual variances. In *Proc. of the 54th IEEE Conference on Decision and Control (CDC'15)*, 2015.
- [8] INRIX official website. Website: <http://www.inrix.com/xd-traffic/>.
- [9] M. J. Lighthill and G. B. Whitham. On kinematic waves. ii. a theory of traffic flow on long crowded roads. *Proc. of the Royal Society of London, Series A, Mathematical and Physical Sciences*, 229(1178):317345, 1955.
- [10] C. F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transport. Res. B-Meth.*, 28B(4):269287, 1994.
- [11] C. F. Daganzo. The cell transmission model, part II: network traffic. *Transport. Res. B-Meth.*, 29B(2):79–93, 1995.
- [12] L. Muñoz, X. Sun, R. Horowitz, and L. Alvarez. A piecewise-linearized cell transmission model and parameter calibration methodology. In *Proc. of the 85th Annual Meeting of the Transportation Research Board (TRB)*, Washington D.C., USA, 2006.
- [13] F. Morbidi, L. L. Ojeda, C. Canudas de Wit, and I. Bellicot. A new robust approach for highway traffic density estimation. In *Proc. of the 13th European Control Conference (ECC'14)*, pages 2576–2580, Strasbourg, France, 2014.
- [14] J. W. C. van Lint and S. P. Hoogendoorn. A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. *Computer-Aided Civil and Infrastructure Engineering*, 25:596–612, 2010.
- [15] D.B. Work, O.-P. Tossavainen, Q. Jacobson, and A.M. Bayen. Lagrangian sensing: traffic estimation with mobile devices. In *Proc. of the American Control Conference (ACC'09)*, pages 1536–1543, June 2009.
- [16] W. Deng, H. Lei, and X. Zhou. Traffic state estimation and uncertainty quantification based on heterogeneous data sources: A three detector approach. *Transport. Res. M-Meth.*, 57:132157, Nov. 2013.
- [17] Y. Li, E. Canepa, and C. Claudel. Optimal control of scalar conservation laws using linear/quadratic programming: Application to transportation networks. *IEEE Trans. Control Netw. Syst.*, 1(1):28–39, march 2014.
- [18] C. Canudas de Wit, F. Morbidi, L. L. Ojeda, A. Y. Kibangou, and I. Bellicot. Grenoble traffic lab: An experimental platform for advanced traffic monitoring and forecasting. *IEEE Control Syst. Mag.*, 2015.
- [19] F. L. Hall, B. L. Allen, and M. A. Gunter. Empirical analysis of freeway flow-density relationships. *Transport. Res. A-Gen.*, 20(3):197 – 210, 1986.
- [20] Centre national d’information routière (CNIR). Website: <http://www.bison-fute.gouv.fr/>.
- [21] Grant M. and S. Boyd. Cvx: Matlab software for disciplined convex programming, 2013. version 2.0 beta, <http://cvxr.com/cvx>.
- [22] Grant M. and S. Boyd. Graph implementations for nonsmooth convex programs. In S. Blondel, S. S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control (a tribute to M. Vidyasagar)*, Lecture Notes in Control and Information Sciences. Springer, 2008.